

DOWNSCALING NCC CGCM OUTPUT FOR SEASONAL PRECIPITATION PREDICTION OVER ISLAMABAD – PAKISTAN

Muhammad Afzaal Karori¹ & Peiqun Zhang²

Abstract:

This study investigates prospects of downscaling for seasonal precipitation prediction over Islamabad, Pakistan. In order to achieve this target large scale variables are taken from CGCM, developed by National Climate Center, CMA, China. NCEP Reanalysis dataset has also been taken into account for comparison. The correlation pattern between observed summer season (JJA) precipitation, over Islamabad, and each variable of CGCM has been obtained. The areas in the pattern, having correlation coefficient $0.4 < R < -0.4$, were identified as predictor windows. Three large scale variables H500, T850 and SLP have been chosen as predictors, after comparing the correlation patterns, for downscaling.

Two statistical downscaling models are used for seasonal precipitation prediction: 1- Coupled Pattern Projection Model (CPPM), in which coupled patterns have been achieved from the covariance between each predictor and predictand and projected onto the predictor field and then calculating the value of predictand on local scale using single variable predictor, 2- Multiple Linear Regression Model (MLRM), in which more than one predictor is used to have predicted time series. Twenty three years (1983 – 2006) observed precipitation data have been used to test these models. Procedures have been repeated for a large number of cases with different size and location of predictor window. CPPM showed good results for forecast verification period i.e. 2003 – 2005. Predicted values were close to reality in this period, but it yielded less agreement between observed and predicted time series for training period i.e. 1983 – 2006. The results of MLRM were found to be very good and were in well accordance with the observed time series of seasonal precipitation, with some exceptions of downscaled values of precipitation which have big difference with observed values. These outlying values may be reduced by further deep investigation in the input of NCC CGCM and the skill of statistical model in representing CGCM output more accurately.

The optimum size and location of predictor window is found to lie over South Pacific Ocean for the predictors H500 and SLP, and over southwest Arabian Sea for T850, which yielded optimum result for seasonal precipitation prediction over Islamabad. Apparently, the predictor area over South Pacific Ocean does not have any thing to do with the rainfall over Islamabad but this area proved to be good for seasonal precipitation prediction when downscaling NCC CGCM output. This aspect of NCC

¹ Pakistan Meteorological Department.

² Climate System Diagnostics and Prediction Division, Beijing Climate Center, CMA, China.

CGCM also requires deep study in its internal structure to improve its ability to regenerate the global circulations close to reality.

Introduction:

Global Circulation Models (GCM) are widely in use in order to have assessment of future climate and climate change scenarios of an area. Although General Circulation Models represent the main features of the global atmospheric circulation reasonably well, their performance in reproducing regional climatic details is rather poor. The resolution of these models usually lies in the range of 150 – 300 kms. The output of these models, at this coarse resolution, does not provide the realistic information on local and regional scales. Particularly, the regional precipitation in the models is not yielded well, because precipitation is influenced by subgrid-scale processes, such as precipitation schemes, physiographical features like topography may not be incorporated cent percent according to reality.

As a result, there is a need to improve some methods to extract GCM's output to have detailed climate information at local and regional scale. These methods are known as downscaling methods. The downscaling can be divided into two groups; Dynamical and Statistical (Murphy et al. 1999). A combination of the above two methods is called as Dynamical-Statistical Downscaling Method. A number of dynamical and statistical techniques have been developed so far which are discussed in (Giorgi et al. 2001). Results from modeling research, global and regional, indicate that both dynamical modeling and statistical downscaling add value to the output from the global models (I. Hanssen-Bauer et al. 2005). These techniques are based on the assumption that large scale variables such as geopotential height, temperatures, sea level pressure etc. are well predicted by GCM.

In this study, Dynamical-statistical downscaling technique has been used. Statistical downscaling is a method to infer local climate information by developing some empirical statistical relationship between large scale fields and local climate conditions (Giorgi et al. 2001). An advantage of statistical downscaling is that it can deal with a variety of scales starting from the station level scale. On the other hand dynamic downscaling is constrained by its resolution. The other advantage of statistical downscaling is that it usually requires less computer time than regional dynamical modeling and it is therefore possible that one can generate a long time series and search out the differences in members of ensemble output (Chen & Li 2004).

In this study, multiple linear regression model and coupled pattern projection model has been used to have seasonal climate prediction of Islamabad, the capital city of Pakistan, with geographical location at 33.6oN & 73.1oE. This technique will be tested for all meteorological stations of Pakistan in future. The physiographical conditions in Pakistan are very much complex (Fig. 1) and cause high variability in local climates over short distances, e.g. from plain areas to mountain regions and from the coast to the inland. The data records of more than forty years are available for most of the meteorological stations in Pakistan, which also provides a good base for the studies in downscaling. This technique is helpful in studying the impacts of climate change, in Pakistan on forestry, agriculture, water resources and many other areas.

Methodology:

Data Used:

Real time observational data are needed for the development and validation of downscaling techniques. Twenty-three years, 1983 – 2006, real time observational seasonal precipitation data of summer (JJA) of Islamabad are used in this study. This dataset was provided by Pakistan Meteorological Department. The time series of precipitation data are then needed to be correlated with the large scale variables.

Large Scale Climate Variables:

The large scale meteorological fields were taken from Global Atmosphere-Ocean Coupled Model (CGCM), which is developed by Beijing Climate Center, China Meteorological Administration, China, and from NCEP Reanalysis data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, USA, from their Web site at <http://www.cdc.noaa.gov> for comparison.

Global Atmosphere-Ocean Coupled Model for Seasonal Prediction:

The global coupled ocean-atmosphere model currently used in the NCC is composed of an AGCM (Version 1.0) named T63L16 AGCM_1.0) which was developed on the basis of the operational medium-range prediction model of the National Meteorological Center of China Meteorological Administration (NMC/CMA), and an OGCM Version 1.0, named GT63L30 OGCM_1.0 which was developed by the State Key Laboratory of Numerical Modeling for Atmospheric Science and Geophysical Fluid Dynamics (LASG) of the IAP/CAS on the basis of the original LASG OGCM (20 levels in vertical direction and $4^\circ \times 5^\circ$ horizontal resolution). These are coupled through the coupling scheme of the Daily Flux Anomaly on the open sea surface. The brief of the whole CGCM model system is shown in Table 1. This model contains large scale variables like SLP, SST, Precipitation and Temperatures at 2M at the resolution $1.875^\circ \times 1.875^\circ$, and H500, H200, T850, U200, V200, U850, V850 at the resolution $2.5^\circ \times 2.5^\circ$.

Seasonal precipitation of summer (JJA), over Islamabad, is taken as Predictand in this study. Predictand is the meteorological element which is needed to be predicted at local scale by downscaling large scale fields from GCMs, the predictors. Selection of predictors is very important to have optimum prediction for a station. A large scale variable should satisfy some conditions (Giorgi et al. 2001) to become a good predictor.

- It is important for a large scale predictor that it should be reproduced by the global or regional climate model, which is in use, realistically.
- They should represent most part of the variability in the predictands.
- The links between predictors and predictands should be strong and stable in time.

The large scale variables used here are Sea Level Pressure, Geopotential Height at 500hPa and Temperatures at 850hPa level. These were chosen as predictors because of having good correlations with the real time precipitation of Islamabad (Figs. 2 – 4). After having the correlation between observed and large scale variables, the areas with high correlation ($0.4 < R < -0.4$) were identified by comparing the linear correlations of observed precipitation with CGCM and with NCEP Reanalysis. These areas are also mentioned within the rectangles in Figs. 2 – 4. These areas contain a few grid points of predictor field and are different for different predictors. Then the area average of temporal covariance between predictand and predictor was calculated in these predictor windows.

Statistical Model:

Multiple Linear Regression Model (MLRM) tend to produce the seasonal climate prediction with little understanding of the physical processes which make the linear relationship between predictor and predictand, while Canonical Correlation Analysis have been utilized for this purpose with greater understanding of underlying physical processes (Lee 2003). However, MLRM has more skill than CCA in terms of regional climate prediction (Lee et al. 1999). In this study, Multi-lead statistical prediction system is used which is a combination of Coupled Pattern Projection Method (CPPM) and MLRM to have optimum seasonal prediction. It was developed by (Lee 2003). This model lies between MLRM and CCA, and development detail is given in (Lee 2003).

The coupled pattern projection model is based on the large scale model variables correlated to a local observed variable, in the study the precipitation. The realization of the predictand is generated from the covariance of predictand and predictor field. When the optimum covariance pattern is determined then the local climate can be predicted by a linear combination of predictors obtained by projecting covariance patterns to the dynamical model prediction data. CPPM consists of two procedures; one is train-forecast procedure, and the other is forecast optimizing procedure.

A time series of coupled mode between predictand and predictor was obtained using following equation,

$$Xp(t) = \mu_y \sum_{i,j} \frac{\text{cov}(i, j) \bullet X(i, j, t)}{\mu_x^2(i, j)} \dots\dots\dots (2.1)$$

where, μ_y is a standard deviation of predictand, $\text{cov}(i, j)$ is covariance pattern between predictor and predictand, $x(i, j, t)$ is the predictor field at time t and μ_x is the standard deviation of predictor.

The transfer function from large scale fields to regional climate, used in this study, is given below,

$$Y_F(t) = \alpha X_p(t) + \beta \dots\dots\dots (2.2)$$

where, $Y_F(t)$ is forecasted time series of predictand, α is regression coefficient and β regression sum of squares, while $X_p(t)$ is obtained from equation (2.1).

The training procedure consists of three steps, which are shown systematically in Fig. 5 (Lee 2003). In first step, a coupled pattern in terms of covariance between predictand and predictor field was obtained using twenty years (1983 – 2002) training period. In the second step, a time series of predictor was generated by projection of coupled pattern onto predictor field for the training period. In the last step, a linear relationship was constructed, using simple linear regression method, between predictand and predictor field. The forecast of the predictand was obtained by applying predictor time series at target years (2003 – 2006), in this study.

To optimize forecast skill, a large number of cases of predictor field with different location and window size were repeated and then multiple linear regressions were used. The equation used for this purpose is given below,

$$Y_F(t) = \mu_y \left[\alpha_o + \sum_{i=1}^N \alpha_i \frac{X_i(t) - \overline{X_i}}{\alpha_{X_i}} \right] \dots\dots\dots (2.3)$$

where, $Y_F(t)$ is a forecasted time series of predictand, μ_y is standard deviation of predictor, α_o is the regression sum of squares, N is total number of independent large scale variables, α_i is the regression coefficient of i th variable, X_i is the predictor time series which is obtained in equation (2.1). Numerator term is variance of projected time series of i th variable which is divided by the standard deviation of the same variable.

Results and Discussion:

Before downscaling, first we should see that the global circulation model is capable of simulating the large scale seasonal circulations of the atmosphere over the area under study. NCC CGCM simulates the monsoonal circulation over Pakistan and surrounding areas with little deviation from the reality. The average pattern of sea level pressure and stream flow at 850hPa over the area during summer season is show in Figs. 6a,b & 7a,b respectively. These figures display the comparisons of CGCM output and NCEP Reanalysis data. It can be seen that the average patterns are almost same with difference of the prominence of monsoonal trough which is more prominent and placed at the position which is close to reality in case of NCEP Reanalysis data (Figs. 7b). Figs. 2 – 4, display the coupled pattern of JJA observed rainfall over Islamabad and CGCM output, colored shades show the correlation pattern and white contours represent the covariance pattern. It can be seen that value of correlation coefficient R is higher in the areas which lie farther from the region under study, while these values are lesser in the areas near to the study region. A large number of covariance patterns over different areas, having correlation coefficient ($0.4 < R < -0.4$), have been obtained and tested using the statistical models aforementioned. Comparison of different time series of seasonal precipitation with observed time series showed different simultaneity. Only the optimum results are included in this report, which show the maximum simultaneity with observed rainfall, especially in the forecast verification period i.e. years 2003 – 2006.

Figs. 8 – 10, represent the comparison of seasonal precipitation obtained from downscaling the CGCM output using single predictor variable, values taken from the

windows shown in Figs. 2 – 4, and observed seasonal precipitation. In Fig. 8, comparison is shown for predictor variable SLP. It can be seen that the predicted precipitation is close to the observed precipitation in forecast verification period, but for training period i.e. year 1983 – 2002, the predicted time series does not match enough with the observed one. The value of correlation coefficient is 0.41. It can not predict extreme points of the observed time series. Fig. 9 shows the comparison of predicted precipitation from large scale variable H500. It can be seen that the fluctuations in the predicted precipitation, in the forecast verification period are according to the observed precipitation but values are not close enough to the reality. The correlation coefficient is 0.53 in this case. The result is really good for the years from 1989 to 1993, but it also does not give a good result for extreme values of observed precipitation. Fig. 10 shows the comparison of predicted precipitation from predictor T850. It shows good prediction result for forecast verification period, but results are not good for training period and the correlation coefficient for this predictor is 0.38.

It revealed from these figures that only one predictor is not enough to have optimum forecast skill for the model; therefore multiple linear regression model (equation 2.3) is used to see the combined effect of all the three predictors to try to have maximum accuracy in prediction. Fig. 11 represents the results of MLRM, and it can be seen most of the points lie close to the observed values with some exceptions. For example, in the year 1996 the predicted value of precipitation is much less than that of observed one and there are also some values which lie close to observed values but with opposite sign. This may be attributed to the calculations of predictors by CGCM. The differences in the downscaled variables mainly represent the differences in natural variability of simulated climate and the skill of statistical models (Chen et al. 2005).

The size and location of predictor window, chosen in this study, lies over south Pacific Ocean for the predictors H500 and SLP, and over southwest Arabian Sea for T850, which yielded optimum result for seasonal precipitation prediction over Islamabad. Positive correlation with temperatures at 850 hPa over southwest Arabia Sea also has some meanings in terms of physical correlation with Islamabad precipitation, because of the south west flow prevails over the area throughout the season (Fig. 7b) and it brings moisture from Arabian Sea. Higher temperatures at 850hPa over this region mean more warm air and more moisture available for precipitation.

Conclusion:

The correlation coefficient between observed and predicted precipitation by the statistical downscaling of NCC CGCM output is 0.65, which means that the regenerated precipitation using MLRM for the years 1983-2006 is in well accordance with the observed precipitation and performance of the prediction system is enhanced when we take more than one predictor. The statistical model performed well in terms of regenerating the precipitation at station/regional scale from large scale variable. This technique is a good tool for seasonal precipitation prediction. However, some outliers exist in the predicted time series. The data values which have opposite sign are considered as outliers. These may have arisen from the input of CGCM or from the misrepresentation of CGCM output by statistical model. The statistical models are not

perfect and rely on the predictors simulated by dynamical models. The outlying values may be reduced by further deep investigation in the input of CGCM and the skill of statistical model in representing CGCM output more accurately.

The location of predictor window also has importance in terms of physical relationship between predictor and predictand. Apparently, the predictor area over South Pacific Ocean does not have any thing to do with the rainfall over Islamabad but this area proved to be good for seasonal precipitation prediction when downscaling NCC CGCM output. This may be attributed to the fact that global circulations simulated by NCC CGCM are not representing the actual situation over the area under study. Deep investigation is required to study NCC CGCM's internal structure to improve its ability to simulate global circulations realistically over the area of study i.e. Pakistan and surrounding areas, or if there is some teleconnection of predictors with this area which is affecting the regional climate of Pakistan.

References:

- Chen D, Li X (2004):** Scale dependent relationship between maximum ice extent in the Baltic Sea and atmospheric circulation. *Global Planet Change*, 41: 275–283
- Chen et al., 2005:** Using Statistical Downscaling to Quantify the GCM-Related Uncertainty in Regional Climate Change Scenarios: A Case Study of Swedish Precipitation. *Advances in Atmospheric Sciences*. 23, 54 – 60.
- Giorgi, F. et al., 2001:** Regional climate simulation—Evaluation and projections. *Climate Change 2001: The Scientific Basis*, J. T. Houghton et al., Eds., Cambridge University Press, 944;thpp.
- I. Hanssen-Bauer et al., 2005:** Statistical downscaling of climate scenarios over Scandinavia. *Clim Res.*, 29, 255 – 268.
- Lee. June-Yi, 2003:** Data and Model Description, In: *Assessment of Potential Seasonal Predictability with a Multi-Model Dynamical-Statistical Ensemble System*. Ph.D Dissertation, Faculty of the Graduate School of the Seoul National University, p 29 - 35.
- Murphy, J., 1999:** An evaluation of statistical and dynamical techniques for downscaling local climate. *J. Climate.*, 12, 2256–2284.

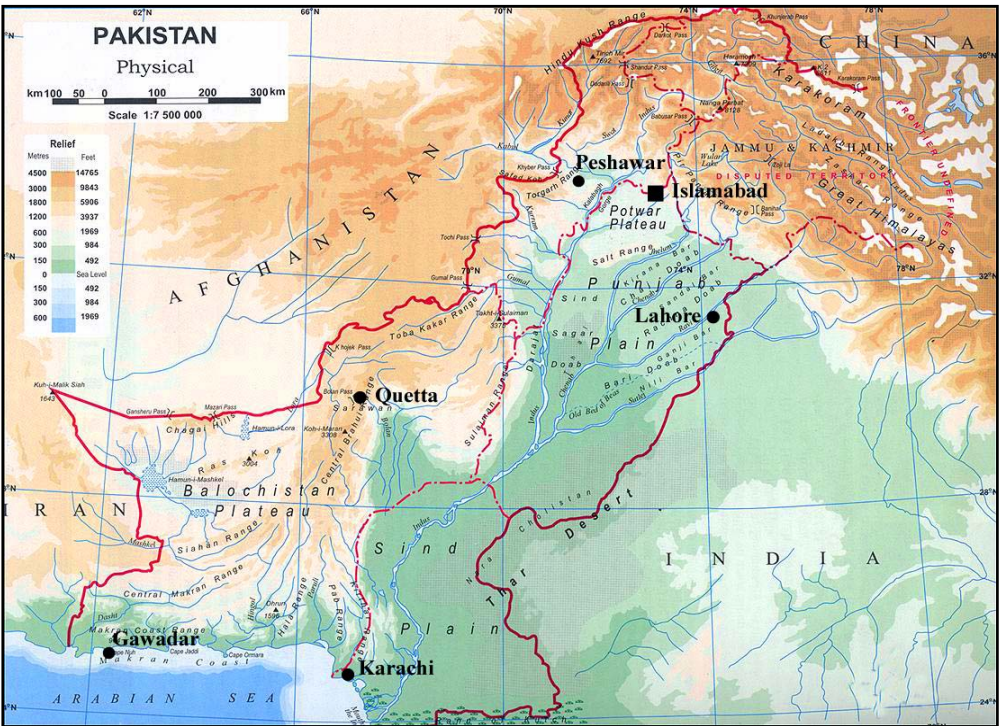


Figure 1: Topographical Features of Pakistan

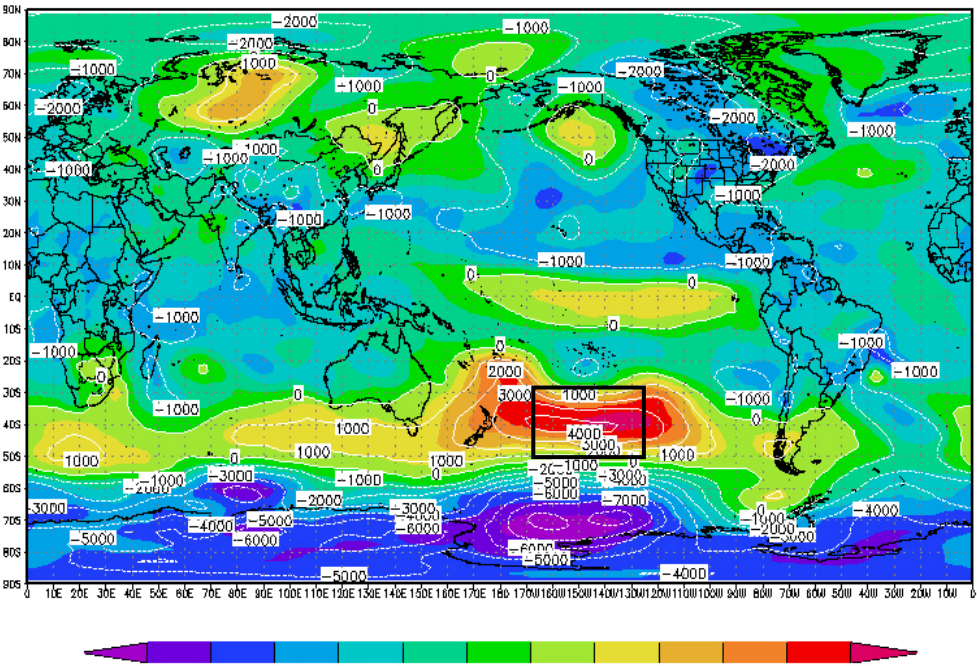
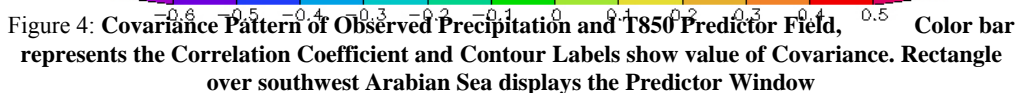
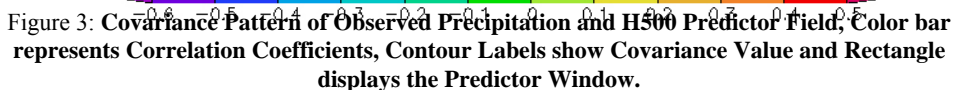


Figure 2: Covariance Pattern of Observed Precipitation and SLP Predictor Field, Color bar represents Correlation Coefficients, Contour Labels show Covariance Value and Rectangle displays the Predictor Window



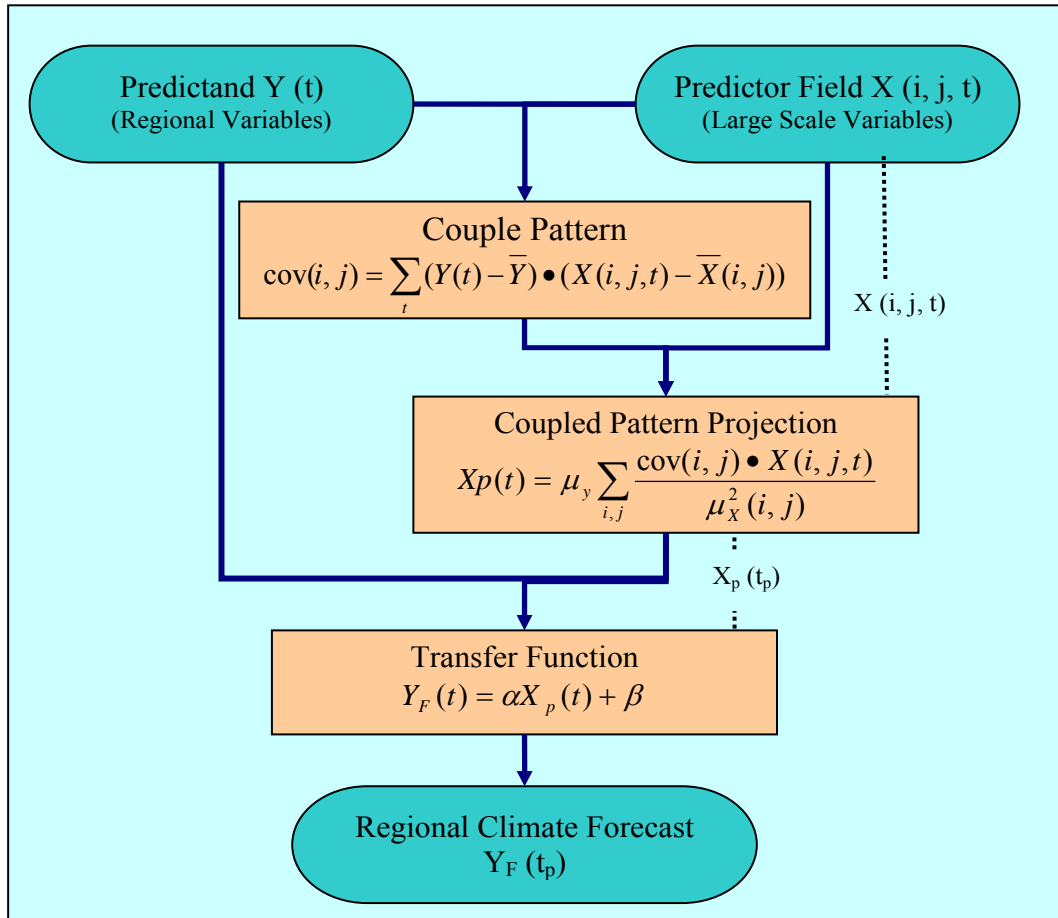


Figure 5: The Schematic Diagram of Prediction Procedure of Coupled Pattern Projection Model (Lee 2003)

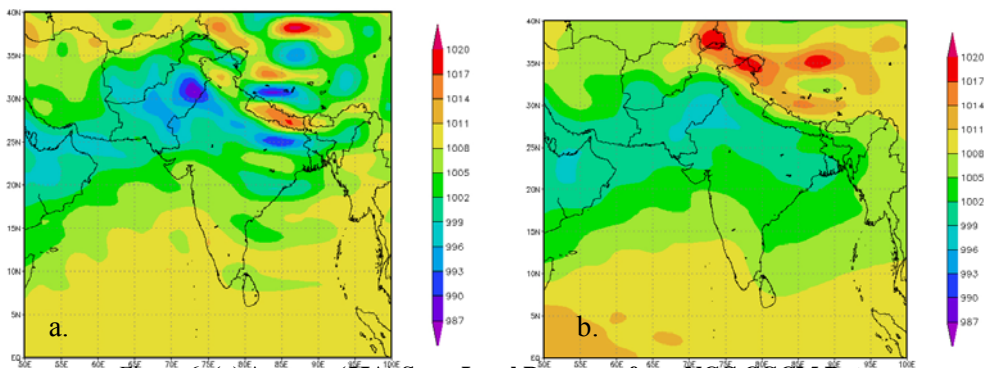


Figure 6: (a) Average (JJA) Sea – Level Pressures from NCC CGCM Data
(b) Average (JJA) Sea – Level Pressures from NCEP Reanalysis Data

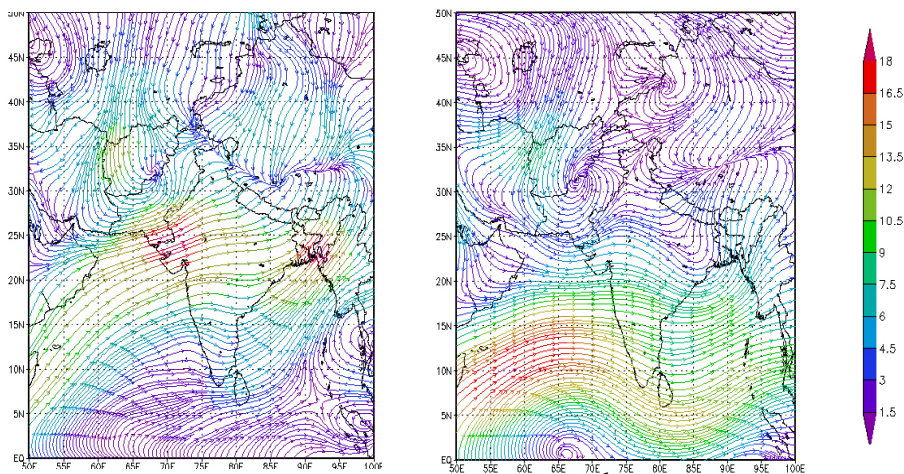


Figure 7: (a) Average (JJA) Stream Flow Pattern at 850hPa from NCC CGCM Data
(b) Average (JJA) Stream Flow Pattern at 850hPa from NCEP Reanalysis

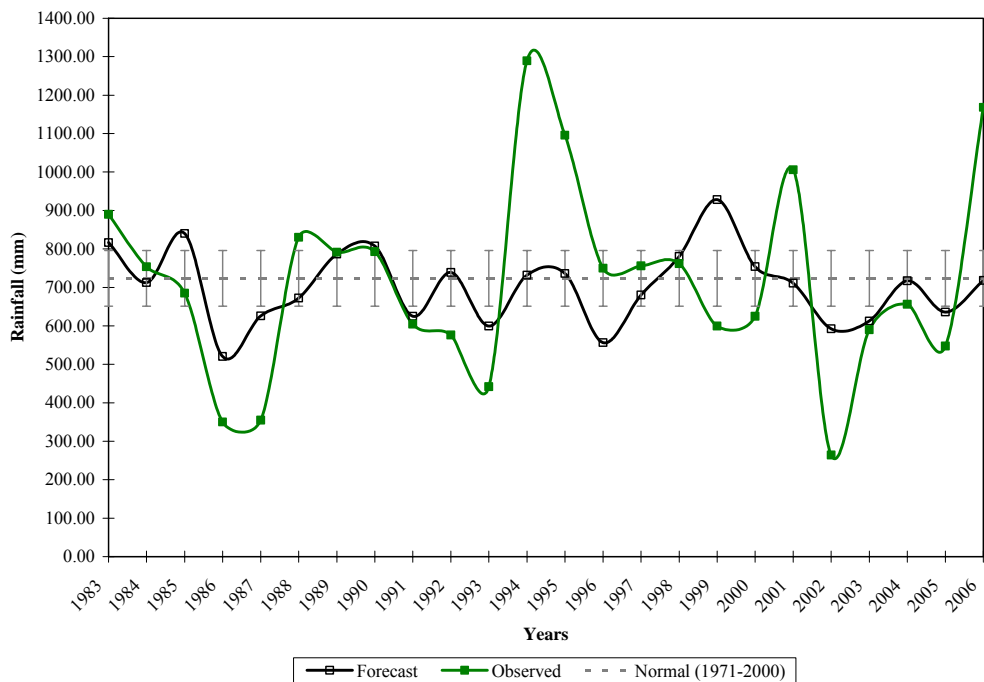


Figure 8: Comparison of Predicted Precipitation with Observed, obtained from single predictor i.e. Sea – Level Pressure Field ($R^2 = 0.41$)

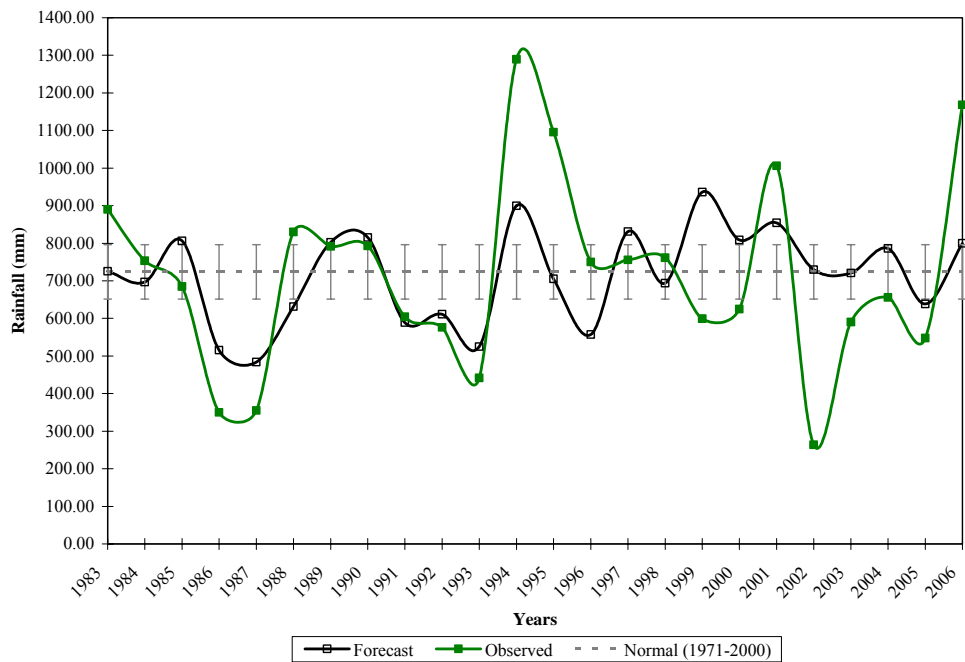


Figure 9: Comparison of Predicted Precipitation with Observed, obtained from single predictor i.e. Geopotential Height at 500hPa Level ($R^2 = 0.53$)

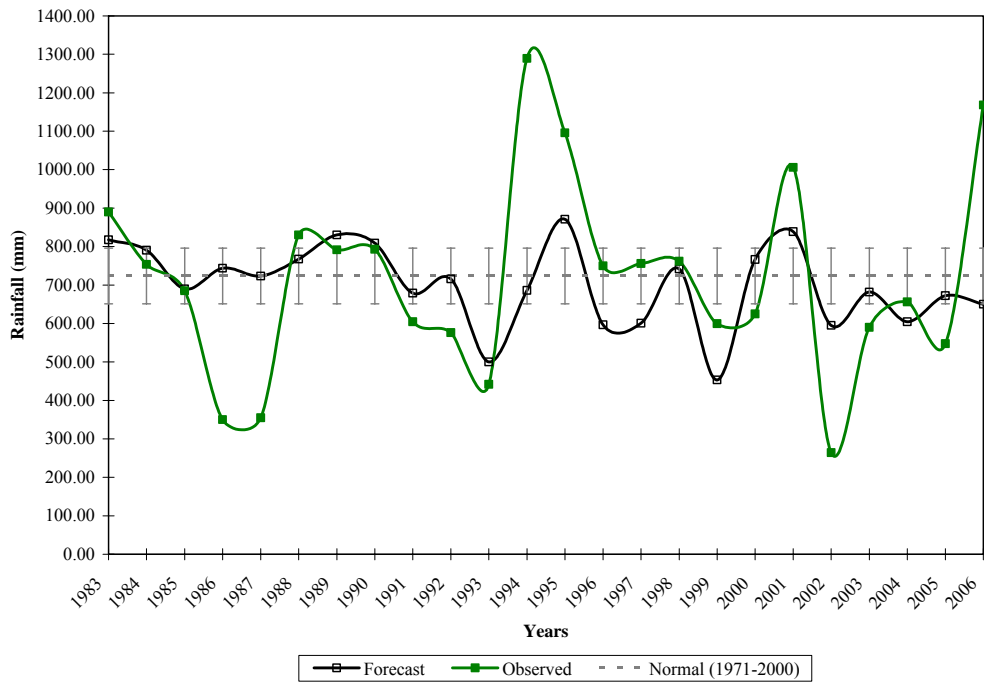


Figure 10: Comparison of Predicted Precipitation with Observed, obtained from single predictor i.e. Temperatures at 850hPa Level ($R^2 = 0.38$)

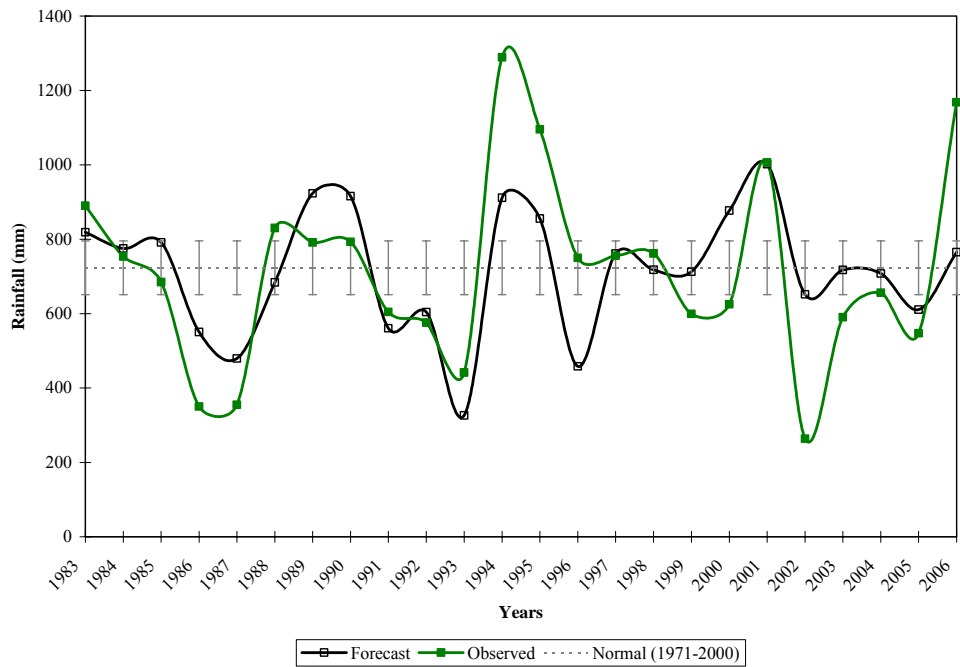


Figure 11: Comparison of Predicted Precipitation with Observed, obtained from MLRM with three predictors i.e. SLP, H500 and T850 ($R^2 = 0.65$)

Table 1: The brief of NCC Global Atmosphere-Ocean Coupled Model (CGCM)

ITEM	CONTENT
Atmospheric component	T63L16
Horizontal resolution	T63 triangle truncation in the horizontal direction with 63 waves, approximately $1.875^{\circ} \times 1.875^{\circ}$
Vertical levels	16 levels P-s hybrid h coordinate with about 25 hPa at the top
Physical processes	Large-scale topography, radiation, large-scale precipitation, cumulus convection, evaporation
Integration scheme	Semi-implicit with 22.5 min timestep
Main modification and development	<p>Introduces a reference atmosphere and mass conservation schemes</p> <p>Uses a stepwise and circular revised method to treat the issue of negative water vapor existing in the initial condition</p> <p>Employs a semi-Lagrangian method to calculate the horizontal and vertical transport of moisture so as to eliminate the occurrence of negative water vapor and pseudo precipitation</p> <p>Uses a static deduction method to reduce modeling truncation errors and improve calculation accuracy of horizontal pressure gradient force near steep topography</p>
Initial condition of land surface	Climatology
Oceanic component	GT63L30
Horizontal resolution	Grid at approximately $1.875^{\circ} \times 1.875^{\circ}$ resolution same as the AGCM
Vertical levels	30 levels h coordinate system with 10 layers in the upper 250m, and 10 layers between 250m to 1000m, 10 layers from 1000m down to 5600m
Governing equations	Baroclinic primitive equation set
Oceanic surface boundary condition	A kind of Newtonian relaxation boundary condition
Integration scheme	<p>An algorithm for separating and coupling barotropic mode and baroclinic mode</p> <p>Inside the baroclinic mode, the thermohaline process is separated from the momentum process</p> <p>All the time integrations of the barotropic mode, the baroclinic mode and the thermohaline process employ leap-frog scheme</p> <p>The time step lengths of the barotropic mode, the baroclinic mode and the thermohaline process are 2 minutes, 4 hours, and 8 hours, respectively</p>
Physical parameterizations	<p>The incorporation of Gent and McWilliams (1990) parameterization scheme of isopycnal surface mixture in order to improve the simulation of the main oceanic thermocline</p> <p>the adoption of Pakanowski and Philander (1981) scheme of vertical mixture of tropical oceanic upper layer between 30°N and 30°S in order to reduce, the horizontal viscosity and improve the simulation of the equatorial thermocline the consideration of penetration of short-wave radiation into oceanic sub-surface layer (Rosati, 1988)</p>
Sea Ice	A thermal sea ice model reconstructed on the basis of the theory of Parkinson and Washington (1979) is used, in which is taken into account the variation of thickness of sea ice, the growth and decline of the areas of sea ice.
Coupling Scheme	Daily Flux Anomaly (DFA) on the open sea surface
Coupling time step	1 day
Initial perturbation	Lagged Average Forecast (LAF) method
Forecast frequency	Twice / yr, in Mar (for Summer) and Oct (for winter and next spring and summer) From 2003, Planned 4 times / yr, in February for Spring, May for Summer, August for Fall, and November for Winter